

Datenmassen verarbeiten

Die Zahlen liegen außerhalb des Vorstellungsbereichs: Täglich werden 2,5 Quintillionen Bytes an Daten produziert – das ist die 2,5 mit 29 Nullen. Die Daten liegen in unterschiedlichsten Formen und Formaten vor. Sie kommen aus Videos und Blogs, aus Fotos, GPS-Signalen von Handys sowie vielen anderen Quellen. Mit Big Data sollen sie miteinander verbunden und nutzbar gemacht werden.

und Velocity (Geschwindigkeit). Diese vier Aspekte laufen parallel und mit gleicher Wertigkeit in die jeweiligen Analysen ein. Bei der Masse an Daten – es geht um Größen, die im Tera-, Peta- oder Zettabytes-Bereich liegen – liegt die Herausforderung darin, diese zu analysieren. Die Vielfalt beschreibt die unterschiedlichen Quellen, aus denen die Daten kommen, und deren Analyse. Dabei wird zwischen strukturierten und unstrukturierten Daten unterschieden. Strukturierte Daten sind bereits aufbereitet und nach festgelegten Attributen geordnet. Sie liegen beispielsweise in Datenbanken vor. Unstrukturierte Daten hingegen beinhalten unterschiedlichste Informationen in ungeordneter Form. Zu finden sind sie zum Beispiel in Bildern und Videos. Um aus ihnen das jeweils Entscheidende herauszufiltern, werden intelligente Verfahren benötigt, die Relevantes von Unrelevantem trennen. Eng damit verbunden ist der Aspekt der Wahrhaftigkeit: Kann den gesammelten Daten vertraut werden, aus welchen Quellen stammen sie? Und schließlich geht es um die Geschwindigkeit der Datenauswertung. In Stockholm beispielsweise können in jeder Sekunde 120.000 GPS-Daten ausgewertet werden.

Um das tägliche Verkehrschaos in Schwedens Hauptstadt Stockholm zu reduzieren, sammelt man in der Stadt ständig Daten aus vielen unterschiedlichen Quellen: zum Beispiel aus den GPS-Daten von Taxis, der Bewegung von Mobiltelefonen und von in der Stadt installierten Kameras. Diese werden in Echtzeit ausgewertet. Innerhalb von Sekunden und automatisch lösen sie Maßnahmen aus, die den Verkehr im Fluss halten. Sollte das System beispielsweise einen Stau erkennen, ändert es von sich aus Ampelschaltungen oder leitet den Verkehr über weniger frequentierte Straßen um. Und dies mit Erfolg: Der Verkehr in Stockholm nahm um 20 Prozent ab, es wurden 50 Prozent kürzere Fahrtzeiten festgestellt, und die Emissionen konnten um 20 Prozent gesenkt werden.

Zusammengefasst führen die „4 Vs“ zu folgender Ereigniskette: Durch die Analyse einer Vielzahl von Daten kommt man zu einer besseren Informationsbasis. Daraus lassen sich bessere Entscheidungen treffen. Und die haben bekanntlich noch nie jemandem geschadet.

Da es beim Thema Big Data um mehr als die im Begriff suggerierten Datenmassen geht, spricht man auch von den „4 Vs“: Volume (Masse), Variety (Vielfalt), Veracity (Wahrhaftigkeit)

Von **Stephan Reimann**,
IBM Information Management Team